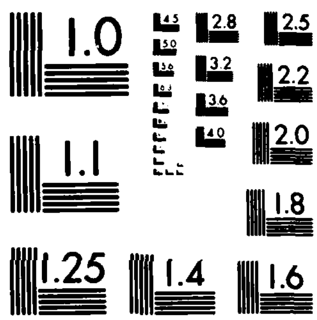
AD-A129 359 PREDICTION OF FUTURE OBSERVATIONS IN POLYNOMIAL GROWTH CURVE MODELS PART 1..(U) PITTSBURGH UNIV PA CENTER FOR MULTIVARIATE ANALYSIS C R RAO MAR 83 83-05-PT-1 1/1

UNCLASSIFIED AFOSR-TR-83-0491-PT-1 F49620-82-K-0001 F/G 12/1 NL

END
DATE
FILMED
7 83
DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963 A

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

2

AD A129359

DTIC FILE COPY

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFOSR-TR- 83-0491	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) PREDICTION OF FUTURE OBSERVATIONS IN POLYNOMIAL GROWTH CURVE MODELS		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL 6. PERFORMING ORG. REPORT NUMBER 83-05
7. AUTHOR(s) C. Radhakrishna Rao		8. CONTRACT OR GRANT NUMBER(s) F49620-82-K-0001
9. PERFORMING ORGANIZATION NAME AND ADDRESS Center for Multivariate Analysis University of Pittsburgh, Ninth Floor, Schenley Hall, Pittsburgh PA 15260		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS PE61102F; 2304/A5
11. CONTROLLING OFFICE NAME AND ADDRESS Mathematical & Information Sciences Air Force Office of Scientific Research Bolling AFB DC 20332		12. REPORT DATE MAR 83 13. NUMBER OF PAGES 13
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED 15a. DECLASSIFICATION DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Empirical Bayes procedure; Compound decision problem; Growth curves; James-Stein estimators; Ridge regression.		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) The problem considered is that of simultaneous prediction of future measurements on a given number of individuals using their past measurements. Assuming a polynomial growth curve model, a number of methods are proposed and their relative efficiencies in terms of the compound mean square prediction error (CMSPE) are compared. There is a similarity between the problem of simultaneous estimation of parameters as considered by Stein and that of simultaneous prediction of future observations. It is found that the empirical Bayes predictor (EBP) based on the empirical Bayes (CONTINUED)		

DTIC ELECTE JUN 13 1983 S D

D

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

83 06 10 029

ITEM #20, CONTINUED: estimator (EBE) of the unknown vector parameters in several linear models proposed by the author (Rao, 1975) has the best possible efficiency compared to the others studied. The problem of determining the appropriate degree of the polynomial growth curve is also studied from the point of view of minimising the CMSPE.

Accession For	
NTIS GRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A	

AFOSR-TR- 83-0491

PREDICTION OF FUTURE OBSERVATIONS IN POLYNOMIAL GROWTH CURVE MODELS PART - 1

C. Radhakrishna Rao

University of Pittsburgh

March 1983

Technical Report No. 83-05

Center for Multivariate Analysis
University of Pittsburgh
Ninth Floor, Schenley Hall
Pittsburgh, PA 15260

This work is sponsored by the Air Force Office of Scientific Research under Contract [redacted]. Reproduction in whole or in part is permitted for any purpose of the United States Government.

F49620-82-K-0001

83 06 10 029

PREDICTION OF FUTURE OBSERVATIONS
IN POLYNOMIAL GROWTH CURVE MODELS
PART - 1*
by

C. Radhakrishna Rao
University of Pittsburgh, Pittsburgh

ABSTRACT

The problem considered is that of simultaneous prediction of future measurements on a given number of individuals using their past measurements. Assuming a polynomial growth curve model, a number of methods are proposed and their relative efficiencies in terms of the compound mean square prediction error (CMSPE) are compared. There is a similarity between the problem of simultaneous estimation of parameters as considered by Stein and that of simultaneous prediction of future observations. It is found that the empirical Bayes predictor (EBP) based on the empirical Bayes estimator (EBE) of the unknown vector parameters in several linear models proposed by the author (Rao, 1975) has the best possible efficiency compared to the others studied. The problem of determining the appropriate degree of the polynomial growth curve is also studied from the point of view of minimising the CMSPE.

AMS Classification: 62C12, 62J07.

Key Words: Empirical Bayes procedure, Compound decision problem, Growth curves, James-Stein estimators, Ridge regression.

*The paper is based on a talk given at the Indian Statistical Institute, Calcutta in December 1981, during the Golden Jubilee celebrations.

1. INTRODUCTION

Let

$$\left.\begin{aligned} \underset{\sim}{Y}_i &= \underset{\sim}{X}\underset{\sim}{\beta}_i + \underset{\sim}{\epsilon}_i \\ y_i &= \underset{\sim}{x}'\underset{\sim}{\beta}_i + \eta_i \end{aligned}\right\} \quad i = 1,\ldots,k \tag{1.1}$$

be k linear models, where $\underset{\sim}{Y}_i$ are observable p-vector random variables, $\underset{\sim}{\beta}_i$ are unknown m-vector parameters, $\underset{\sim}{X}$ is $(p \times m)$ and $\underset{\sim}{x}$ is $(m \times 1)$ given matrices. The problem is to predict $y_1,\ldots,y_k$ given $\underset{\sim}{Y}_1,\ldots,\underset{\sim}{Y}_k$ when the dispersion matrix of the error term $(\underset{\sim}{\epsilon}_i,\eta_i)$ is of the form

$$\sigma^2 \begin{pmatrix} \underset{\sim}{V}_{11} & \underset{\sim}{V}_{12} \\ \underset{\sim}{V}_{21} & \underset{\sim}{V}_{22} \end{pmatrix} = \sigma^2 \underset{\sim}{V} \tag{1.2}$$

where σ^2 is unknown and $\underset{\sim}{V}_{ij}$ are known. We shall assume that $\underset{\sim}{X}$ and $\underset{\sim}{V}_{11}$ are of full rank and $\underset{\sim}{V}_{22} \neq 0$. Suitable modifications can be made when these matrices are deficient in rank (see Rao, 1973, pp. 296-302).

If no assumption is made about the joint distribution of $(\underset{\sim}{Y}_i,y_i)$, then the least squares theory may be applied to estimate $\underset{\sim}{\beta}_i$ and y_i simultaneously. This leads to minimization of

$$\begin{pmatrix} \underset{\sim}{Y}_i - \underset{\sim}{X}\underset{\sim}{\beta}_i \\ y_i - \underset{\sim}{x}'\underset{\sim}{\beta}_i \end{pmatrix}' \begin{pmatrix} \underset{\sim}{V}_{11} & \underset{\sim}{V}_{12} \\ \underset{\sim}{V}_{21} & \underset{\sim}{V}_{22} \end{pmatrix}^{-1} \begin{pmatrix} \underset{\sim}{Y}_i - \underset{\sim}{X}\underset{\sim}{\beta}_i \\ y_i - \underset{\sim}{x}'\underset{\sim}{\beta}_i \end{pmatrix} \tag{1.3}$$

with respect to $\underset{\sim}{\beta}_i$ and y_i. The estimates so obtained are easily seen to be

$$\begin{aligned} \underset{\sim}{\beta}_i^{\ell} &= (\underset{\sim}{X}'\underset{\sim}{V}_{11}^{-1}\underset{\sim}{X})^{-1}\,\underset{\sim}{X}'\underset{\sim}{V}_{11}^{-1}\underset{\sim}{Y}_i \\ \hat{y}_i &= \underset{\sim}{x}'\underset{\sim}{\beta}^{\ell} + \underset{\sim}{V}_{21}\underset{\sim}{V}_{11}^{-1}(\underset{\sim}{Y}_i - \underset{\sim}{X}\underset{\sim}{\beta}_i^{\ell}) \end{aligned} \tag{1.4}$$

where $\underset{\sim}{\beta}_i^{\ell}$ is the least squares estimator of $\underset{\sim}{\beta}_i$ from the model i. The same estimator $\hat{y}_i$ can be deduced by considering a linear function $\underset{\sim}{L}'\underset{\sim}{Y}_i + a$ and minimizing the mean square error

$$E(y_i - \underset{\sim}{L}'\underset{\sim}{Y}_i - a)^2$$

subject to the unbiasedness condition

$$E(y_i - \underset{\sim}{L}'\underset{\sim}{Y}_i - a) = 0. \tag{1.5}$$

Predictors of the type (1.4), which may be called the best linear unbiased predictor (BLUP), have been studied by Rao (1973, p. 234) and Toutenburgh (1970). The MSPE (mean square prediction error) of the BLUP, $\hat{y}_i$ in (1.4), is

$$E(y_i - \hat{y}_i)^2 = \sigma^2(\underset{\sim}{V}_{22} - \underset{\sim}{V}_{21}\underset{\sim}{V}_{11}^{-1}\underset{\sim}{V}_{12} + \underset{\sim}{d}'\underset{\sim}{U}\underset{\sim}{d}) \tag{1.6}$$

where $\underset{\sim}{d} = \underset{\sim}{x} - \underset{\sim}{X}'\underset{\sim}{V}_{11}^{-1}\underset{\sim}{V}_{12}$ and $\underset{\sim}{U} = (\underset{\sim}{X}'\underset{\sim}{V}_{11}^{-1}\underset{\sim}{X})^{-1}$.

In this paper, we examine the possibility of constructing predictors $\hat{y}_i$ of y_i without using the condition (1.5), such that the CMSPE (the compound MSPE)

$$E[\sum_1^k (y_i - \hat{y}_i)^2] \tag{1.7}$$

is a minimum. Such a procedure leads to predictors analogous to Stein type estimators in simultaneous estimation of parameters (see Stein, 1955 and James and Stein, 1961).

2. BEST LINEAR PREDICTORS (BLP)

If $\underset{\sim}{L}'\underset{\sim}{Y}_i + a$ is a linear predictor of y_i, then the CMSPE is

$$\begin{aligned} E\sum_{i=1}^k (y_i - \underset{\sim}{L}'\underset{\sim}{Y}_i - a)^2 &= E\sum_1^k [y_i - \underset{\sim}{x}'\underset{\sim}{\beta}_i - \underset{\sim}{V}_{21}\underset{\sim}{V}_{11}^{-1}(\underset{\sim}{Y}_i - \underset{\sim}{X}\underset{\sim}{\beta}_i)]^2 \\ &\quad + E\sum_1^k [\underset{\sim}{d}'\underset{\sim}{\beta}_i - \underset{\sim}{M}'\underset{\sim}{Y}_i - a]^2 \end{aligned} \tag{2.1}$$

where $\underset{\sim}{d} = \underset{\sim}{x}-\underset{\sim}{X}'\underset{\sim}{V}_{11}^{-1}\underset{\sim}{V}_{12}$ and $\underset{\sim}{M} = \underset{\sim}{L}-\underset{\sim}{V}_{11}^{-1}\underset{\sim}{V}_{12}$. The first term on the right hand side of (2.1) does not involve $\underset{\sim}{L}$ and a, and in order to minimize the CMSPE we need only consider the second term. It is easily seen that the minimum is attained for given $\underset{\sim}{\beta}_1,\ldots,\underset{\sim}{\beta}_k$ when

$$\underset{\sim}{L} = (\underset{\sim}{V}_{11}+\underset{\sim}{X}\underset{\sim}{F}\underset{\sim}{X}')^{-1}(\underset{\sim}{V}_{12}+\underset{\sim}{X}\underset{\sim}{F}\underset{\sim}{x}) \text{ and } a = (\underset{\sim}{x}'-\underset{\sim}{L}'\underset{\sim}{X})\underset{\sim}{\alpha} \tag{2.2}$$

where

$$\underset{\sim}{\alpha} = k^{-1}\,\Sigma\,\underset{\sim}{\beta}_i \quad \text{and} \quad \sigma^2\underset{\sim}{F} = k^{-1}\Sigma(\underset{\sim}{\beta}_i-\underset{\sim}{\alpha})(\underset{\sim}{\beta}_i-\underset{\sim}{\alpha})'. \tag{2.3}$$

The optimum values of $\underset{\sim}{L}$ and a as obtained in (2.2) involve two functions $\underset{\sim}{\alpha}$ and $\underset{\sim}{F}$ of the unknown parameters $\underset{\sim}{\beta}_1,\ldots,\underset{\sim}{\beta}_k$ and σ^2. If $\underset{\sim}{\alpha}$ and $\underset{\sim}{F}$ are known, then the BLP of y_i which minimize the CMSPE are

$$\begin{aligned} &\underset{\sim}{d}'[\underset{\sim}{\beta}_i^{\ell}-\underset{\sim}{U}(\underset{\sim}{F}+\underset{\sim}{U})^{-1}(\underset{\sim}{\beta}_i^{\ell}-\underset{\sim}{\alpha})] + \underset{\sim}{V}_{21}\underset{\sim}{V}_{11}^{-1}\underset{\sim}{Y}_i \\ &\qquad = \underset{\sim}{d}'\underset{\sim}{\beta}_i^{b} + \underset{\sim}{V}_{21}\underset{\sim}{V}_{11}^{-1}\underset{\sim}{Y}_i \quad i=1,\ldots,k, \end{aligned} \tag{2.4}$$

where $\underset{\sim}{U} = (\underset{\sim}{X}'\underset{\sim}{V}_{11}^{-1}\underset{\sim}{X})^{-1}$, $\underset{\sim}{\beta}_i^{\ell}$ is the least squares estimator of $\underset{\sim}{\beta}_i$ and

$$\underset{\sim}{\beta}_i^{b} = \underset{\sim}{\beta}_i^{\ell} - \underset{\sim}{U}(\underset{\sim}{F}+\underset{\sim}{U})^{-1}(\underset{\sim}{\beta}_i^{\ell}-\underset{\sim}{\alpha}). \tag{2.5}$$

It may be noted that when $\underset{\sim}{\alpha}$ and $\underset{\sim}{F}$ are known, $\underset{\sim}{\beta}_i^{b}$ is the best linear estimator of $\underset{\sim}{\beta}_i$, $i = 1,\ldots,k$, in the sense that the matrix

$$E\sum_1^k(\underset{\sim}{b}_i-\underset{\sim}{\beta}_i)(\underset{\sim}{b}_i-\underset{\sim}{\beta}_i)'-E\sum_1^k(\underset{\sim}{\beta}_i^{b}-\underset{\sim}{\beta}_i)(\underset{\sim}{\beta}_i^{b}-\underset{\sim}{\beta}_i)' \tag{2.6}$$

where b_i are any linear estimators of β_i and the expectations are taken for fixed $\beta_1,\ldots,\beta_k$, is non-negative definite. Further, $\underset{\sim}{\beta}_i^{b}$ may also be recognized as the Bayes estimator of $\underset{\sim}{\beta}_i$ when $\underset{\sim}{Y}_i$ has the multivariate normal distribution

and $\boldsymbol{\beta}_i$ has the prior multivariate normal distribution with mean $\boldsymbol{\alpha}$ and variance-covariance matrix $\sigma^2\mathbf{F}$, in the sense that the matrix

$$E(\mathbf{b}_i-\boldsymbol{\beta}_i)(\mathbf{b}_i-\boldsymbol{\beta}_i)'-E(\boldsymbol{\beta}_i^b-\boldsymbol{\beta}_i)(\boldsymbol{\beta}_i^b-\boldsymbol{\beta}_i) \tag{2.7}$$

where the expectations are taken over variations of $\mathbf{Y}_i$ and $\boldsymbol{\beta}_i$, is non-negative definite (see Lindley and Smith, 1972).

The average MSPE for the BLP's in (2.4) is

$$\begin{aligned} k^{-1}\, E \sum_1^k(\mathbf{d}'\boldsymbol{\beta}_i^b+\mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{Y}_i - y_i)^2 &= \sigma^2(\mathbf{V}_{22}-\mathbf{V}_{21}\mathbf{V}_{11}^{-1}\mathbf{V}_{12}+\mathbf{d}'\mathbf{U}\mathbf{d}) \\ &\quad - \sigma^2\mathbf{d}'\mathbf{U}(\mathbf{F}+\mathbf{U})^{-1}\mathbf{U}\mathbf{d}. \end{aligned} \tag{2.8}$$

Comparing the average MSPE's of the BLUP's given in (1.6) and the BLP's given in (2.8), we find that the last term in (2.8) represents the reduction in loss when $\boldsymbol{\alpha}$ and $\mathbf{F}$ are known.

3. EMPIRICAL BAYES PREDICTORS (EBP)

The best linear estimator $\boldsymbol{\beta}_i^b$ of $\boldsymbol{\beta}_i$ as defined in (2.5) involves the knowledge of $\boldsymbol{\alpha}$ and $\mathbf{F}$. If they are not known, we can substitute for them suitable estimates and obtain modified estimators of $\boldsymbol{\beta}_i$. Natural estimators of $\boldsymbol{\alpha}$, σ^2 and $\mathbf{G} = \sigma^2(\mathbf{F}+\mathbf{U})$ are of the form

$$\begin{aligned} \hat{\boldsymbol{\alpha}} &= k^{-1}\Sigma\,\boldsymbol{\beta}_i^\ell \\ \hat{\sigma}^2 &= c_1^{-1}w = c_1^{-1}\sum_1^k(\mathbf{Y}_i'\mathbf{V}_{11}^{-1}\mathbf{Y}_i-\mathbf{Y}_i'\mathbf{V}_{11}^{-1}\mathbf{X}\boldsymbol{\beta}_i^\ell) \\ \hat{\mathbf{G}} &= c_2^{-1}\mathbf{B} = c_2^{-1}\sum_1^k(\boldsymbol{\beta}_i^\ell-\hat{\boldsymbol{\alpha}})(\boldsymbol{\beta}_i^\ell-\hat{\boldsymbol{\alpha}})' \end{aligned}$$

where c_1 and c_2 are constants. Substituting these estimates in (2.5) we obtain

$$\underset{\sim}{\beta}_i^e = \underset{\sim}{\beta}_i^\ell - c\, w \underset{\sim}{U}\underset{\sim}{B}^{-1}(\underset{\sim}{\beta}_i^\ell - \hat{\underset{\sim}{\alpha}}) \tag{3.1}$$

where c is a constant to be suitably chosen. The estimator $\underset{\sim}{\beta}_i^e$ which is no longer linear in Y_i may be called the EBE (empirical Bayes estimator, although the terminology may not be appropriate without introducing an apriori distribution for $\underset{\sim}{\beta}_i$). Estimators of the type (3.1) have been considered by Efron and Morris (1972, 1975) and Rao (1953, 1975).

In the paper (Rao, 1975), the author has shown that when c in (3.1) is chosen as $(k-m-2)/(kp-km+2)$, where m = rank of $\underset{\sim}{X}$, the expectation of the matrix

$$\sum_{i=1}^{k} (\underset{\sim}{\beta}_i^\ell - \underset{\sim}{\beta}_i)(\underset{\sim}{\beta}_i^\ell - \underset{\sim}{\beta}_i)' - \sum_{i=1}^{k} (\underset{\sim}{\beta}_i^e - \underset{\sim}{\beta}_i)(\underset{\sim}{\beta}_i^e - \underset{\sim}{\beta}_i)' \tag{3.2}$$

for any fixed $\beta_1,\ldots,\beta_k$, is non-negative definite, which implies that

$$E[\sum_1^k (\underset{\sim}{d}'\underset{\sim}{\beta}_i^\ell - \underset{\sim}{d}'\underset{\sim}{\beta}_i)^2] \geq E[\sum_1^k (\underset{\sim}{d}'\underset{\sim}{\beta}_i^e - \underset{\sim}{d}'\underset{\sim}{\beta}_i)^2]. \tag{3.3}$$

Now substituting $\underset{\sim}{\beta}_i^e$ for $\underset{\sim}{\beta}_i^b$ in (2.4), we obtain the modified predictor for y_i

$$\tilde{y}_i = \underset{\sim}{d}'\underset{\sim}{\beta}_i^e + \underset{\sim}{V}_{21}\underset{\sim}{V}_{11}^{-1}\underset{\sim}{Y}_i \tag{3.4}$$

which may be called the EBP (empirical Bayes predictor).

Upto now we have not made any distributional assumptions. If we assume a multivariate normal distribution for $(\underset{\sim}{Y}_i, y_i)$, then

$$E(y_i|\underset{\sim}{Y}_i) = \underset{\sim}{d}'\underset{\sim}{\beta}_i + \underset{\sim}{V}_{21}\underset{\sim}{V}_{11}^{-1}\underset{\sim}{Y}_i$$

$$E[\sum_1^k(\tilde{y}_i - y_i)^2] = E\sum_1^k[y_i - E(y_i|\underset{\sim}{Y}_i)]^2 + E\sum_1^k[E(y_i|\underset{\sim}{Y}_i) - \tilde{y}_i]^2$$

$$= k\,\sigma^2(\underset{\sim}{V}_{22} - \underset{\sim}{V}_{21}\underset{\sim}{V}_{11}^{-1}\underset{\sim}{V}_{12}) + E\sum_1^k(\underset{\sim}{d}'\underset{\sim}{\beta}_i^e - \underset{\sim}{d}'\underset{\sim}{\beta})^2. \tag{3.5}$$

The corresponding expression for the BLUP (1.4) is

$$E\sum_1^k(\hat{y}_i-y_i)^2 = k\,\sigma^2(V_{22}-V_{21}V_{11}^{-1}V_{12}) + E\sum_1^k d'\beta_i^{\ell}-d'\beta_i)^2. \tag{3.6}$$

Using the result (3.3), and comparing (3.5) and (3.6), we find that

$$E[\sum_1^k(\tilde{y}_i-y_i)^2] \le E[\sum_1^k(\hat{y}_i-y_i)^2] \tag{3.7}$$

where the expectations are taken for any fixed set $\beta_1,\ldots,\beta_k$ of the true values. Thus the EBP's of $y_1,\ldots,y_k$ are uniformly better than the BLUP's.

4. PREDICTION IN POLYNOMIAL GROWTH CURVE MODELS

Let y_{ti} be the measurement at time t on individual i. We consider the problem of predicting $y_{p+1,i}$ on the basis of $y_{1i},\ldots,y_{pi}$ assuming a polynomial growth curve model

$$y_{ti} = \beta_{0i}\psi_0(t)+\ldots+\beta_{si}\psi_s(t)+\epsilon_{ti} \tag{4.1}$$

$$t=1,\ldots,p+1;\ i=1,\ldots,k$$

where $\psi_0(t),\psi_1(t),\ldots$ are orthogonal polynomials such that

$$\sum_1^p \psi_r(t)\psi_m(t) = 1 \text{ for } r=m \text{ and } 0 \text{ for } r\neq m$$

and

$$V(\epsilon_{ti}) = \sigma^2,\ \operatorname{cov}(\epsilon_{ti},\epsilon_{ui}) = 0,\ t\neq u,$$

$$\operatorname{cov}(\epsilon_{ti},\epsilon_{uj}) = 0,\ i\neq j.$$

Under the model (4.1), the least squares estimators of β_{ri}, σ^2 are

$$\beta_{ri}^{\ell} = \sum_{t=1}^{p} y_{ti}\psi_r(t),\ r=0,1,\ldots,s;\ i=1,\ldots,k, \tag{4.2}$$

$$\hat{\sigma}^2 = [\sum_{1}^{p}\sum_{1}^{k} y_{ti}^2 - \sum_{0}^{s}\sum_{1}^{k}(\beta_{ri}^{\ell})^2] \div k(p-s-1). \tag{4.3}$$

Also, for given i

$$V(\beta^{\ell}_{ri}) = \sigma^2 \quad \text{and} \quad \text{cov}(\beta^{\ell}_{ri}, \beta^{\ell}_{qi}) = 0,\ r \neq q.$$

We consider different methods of predicting $y_{p+1,i}$, $i = 1,\ldots,k$, simultaneously and compare their CMSPE (compound mean square prediction error).

4.1 BLUP with a subset of terms

If we choose only the first (q+1) terms in the model (4.1), then the BLUP of $y_{p+1,i}$ is

$$y^{\ell}_{p+1,i} = \beta^{\ell}_{0i}\psi_0(p+1)+\ldots+\beta^{\ell}_{qi}\psi_q(p+1) \tag{4.1.1}$$

and the MSPE for given i is

$$\sigma^2 \sum_0^q [\psi_r(p+1)]^2 + [\sum_{q+1}^s \beta_{ri}\psi_r(p+1)]^2$$

when in fact the true model has all the (s+1) terms. The corresponding CMSPE is

$$k\sigma^2 \sum_0^q [\psi_r(p+1)]^2 + \sum_{i=1}^k [\sum_{r=q+1}^s \beta_{ri}\psi_r(p+1)]^2. \tag{4.1.2}$$

If all the (s+1) terms in (4.1) are used, then the CMSPE is

$$k\sigma^2 \sum_0^s [\psi_r(p+1)]^2. \tag{4.1.3}$$

The omission of the last s-q terms in (4.1) provides better prediction, although the corresponding regression coefficients may not be zero, if

$$k\sigma^2 \sum_{q+1}^s [\psi_r(p+1)]^2 > \sum_{i=1}^k [\sum_{q+1}^s \beta_{ri}\psi_r(p+1)]^2 \tag{4.1.4}$$

which might hold when the last regression coefficients are small. The best choice of q is that value for which (4.1.2) is a minimum.

In practice, the minimization of (4.1.2) over q cannot be carried out since σ^2 and β_{ri} are not known. An estimate of q may be obtained by minimizing an estimate of (4.1.2), which is

$$k\hat{\sigma}^2 \sum_{0}^{s}[\psi_r(p+1)]^2 + \sum_{i=1}^{k} [\sum_{r=q+1}^{s} \beta_{ri}^{\ell}\psi_r(p+1)]^2 - 2k\hat{\sigma}^2 \sum_{q+1}^{s} [\psi_r(p+1)]^2. \qquad (4.1.5)$$

Since the first term in (4.1.5) does not depend on q, we need only minimize the expression

$$\sum_{i=1}^{k} [\sum_{r=q+1}^{s} \beta_{ri}^{\ell}\psi_r(p+1)]^2 - 2k\hat{\sigma}^2 \sum_{q+1}^{s} [\psi_r(p+1)]^2 \qquad (4.1.6)$$

which is analogous to the criteria suggested by Akaike (1973) and Shibata (1981) in the context of fitting a model to observed data. The emphasis in our case is on the prediction of future observations and the procedures suggested by Akaike and Shibata for obtaining a good fit to observed data may not be appropriate.

Table 1 gives the wieghts of 13 male mice measured at intervals of 3 days over the 21 days from birth to weaning, as reported by Williams and Izenman (1981).

Table 1. Weights of 13 male mice measured at successive intervals of 3 days over 21 days from birth to weaning

mice \ days	3	6	9	12	15	18	21
1	0.109	0.388	0.621	0.823	1.078	1.132	1.191
2	0.218	0.393	0.568	0.729	0.839	0.852	1.004
3	0.211	0.394	0.549	0.700	0.783	0.870	0.925
4	0.209	0.419	0.645	0.850	1.001	1.026	1.069
5	0.193	0.362	0.520	0.530	0.641	0.640*	0.751
6	0.201	0.361	0.502	0.530	0.657	0.762	0.888
7	0.202	0.370	0.498	0.650	0.795	0.858	0.910
8	0.190	0.350	0.510	0.666	0.819	0.879	0.929
9	0.219	0.399	0.578	0.699	0.709	0.822	0.953
10	0.255	0.400	0.545	0.690	0.796	0.825	0.836
11	0.224	0.381	0.577	0.756	0.869	0.929	0.999
12	0.187	0.329	0.441	0.525	0.589	0.621	0.796
13	0.278	0.471	0.606	0.770	0.888	1.001	1.105

*This could be a recording error, but no change was made in the present computations.

In each case, the seventh measurement is predicted using the first six by BLUP, the formula (4.1.1), for different values of q (degree of the polynomial). The sums of squared differences (SSD) between the observed and predicted over the 13 mice for each q were as follows.

q:	0	1	2	3	4	5
SSD:	1.7911	0.2063	0.1042	0.1750	0.5991	7.4700

It is interesting to note that the second degree polynomial provides the best model for predicting the seventh observation, and the predicion becomes worse as we increase the degree of the polynomial, although higher degree polynomials should theoretically provide a better fit to the data.

Since it is found that a second degree polynomial is the appropriate model for predicting future observations, we shall explore alternative methods of estimating the regression coefficients $(\beta_{0i}, \beta_{1i}, \beta_{2i})$ for prediction purposes and examine their predictive efficiencies.

4.2 James-Stein regression predictor (JSRP)

The least squares estimators, $\beta^{\ell}_{0i}, \beta^{\ell}_{1i}, \beta^{\ell}_{2i}$ and $\hat{\sigma}^2$ (with $p=6$, $s=3$, $k=13$), are computed as in formulae (4.2) and (4.3). Using these, the J-S estimators of $\beta_{0i}, \beta_{1i}, \beta_{2i}$ are obtained as follows:

$$\beta^J_{ri} = \left(1 - \frac{26\hat{\sigma}^2}{28S_i^2}\right)\beta^{\ell}_{ri}, \quad S_i^2 = \sum_0^2 (\beta^{\ell}_{ri})^2. \tag{4.2.1}$$

The predictor of y_{7i} based on the estimators (4.2.1) is

$$y^J_{7i} = \beta^J_{0i}\psi_0(p+1) + \beta^J_{1i}\psi_1(p+1) + \beta^J_{2i}\psi_2(p+1). \tag{4.2.2}$$

It is known that the J-S estimators have smaller compound mean square error than the least square estimators, which may not imply that any linear function

of the unknown parameters is better estimated by substituting the J-S estimators instead of the least square estimators (see Rao and Shinozaki, 1978 and Rao, 1980).

4.3 Shrunken regression predictor (SRP)

Let $\bar{\beta}_r = k^{-1}(\beta^{\ell}_{r1}+\ldots+\beta^{\ell}_{rk})$ and $\hat{\sigma}^2$ be as computed in (4.3) with $p=6$, $s=3$ and $k=13$. Then estimate β_{ri} by shrinking the least squares estimators using the formula

$$\beta^{s}_{ri} = \frac{\bar{\beta}^2_r}{\bar{\beta}^2_r+\hat{\sigma}^2}\,\beta^{\ell}_{ri},\quad r=0,1,2. \tag{4.3.1}$$

The predictor of y_{7i} based on (4.3.1) is

$$y^{s}_{7i} = \beta^{s}_{0i}\psi_0(p+1)+\beta^{s}_{1i}\psi_1(p+1)+\beta^{s}_{2i}\psi_2(p+1). \tag{4.3.1}$$

4.4 Empirical Bayes predictor (EBP)

Let

$$\begin{aligned}
\underset{\sim}{\beta}^{\ell}_{i} &= (\beta^{\ell}_{0i},\beta^{\ell}_{1i},\beta^{\ell}_{2i})',\\
\underset{\sim}{\bar{\beta}} &= (\bar{\beta}_0,\bar{\beta}_1,\bar{\beta}_2)',\\
\underset{\sim}{B} &= \sum_{1}^{13}(\underset{\sim}{\beta}^{\ell}_{i}-\underset{\sim}{\bar{\beta}})(\underset{\sim}{\beta}^{\ell}_{i}-\underset{\sim}{\bar{\beta}})'.
\end{aligned}$$

Then the EBE of $\underset{\sim}{\beta}_i = (\beta_{0i},\beta_{1i},\beta_{2i})$ is, as shown in Rao (1975),

$$\underset{\sim}{\beta}^{e}_{i} = \underset{\sim}{\beta}^{\ell}_{i} - \frac{(k-q-3)(p-s-1)\hat{\sigma}^2}{k(p-s-1)+2}\,\underset{\sim}{B}^{-1}(\underset{\sim}{\beta}^{\ell}_{i}-\underset{\sim}{\bar{\beta}}). \tag{4.4.1}$$

The predictor of y_{7i} based on (4.4.1), with $k=13$, $q=2$, $p=6$, $s=3$, is

$$y^{e}_{7i} = \beta^{e}_{0i}\psi_0(p+1) + \beta^{e}_{1i}\psi_1(p+1) + \beta^{e}_{2i}\psi_2(p+1). \tag{4.4.2}$$

It may be noted that β^{e}_{i} is also a Stein type estimator (see Efron and Morris, 1972 and Rao, 1975) of a vector parameter. Since the estimators β^{ℓ}_{ri}, $r=0,1,2$

and $i = 1,\ldots,13$ are all uncorrelated, the problem may also be considered as the simultaneous estimation of the 39 parameters β_{ri}. This was not tried in the present data analysis.

4.5 Ridge Regression predictor

The ridge regression estimator of β_{ri} is computed from the formula (see Hoel and Kennard, 1970)

$$\beta^R_{si} = \frac{S_i^2}{S_i^2+3\hat{\sigma}^2}\,\beta^\ell_{si},\quad S_i^2 = \sum_0^2 (\beta^\ell_{si})^2. \tag{4.5.1}$$

The predictor of y_{7i} based on (4.5.1) is

$$y^R_{7i} = \beta^R_{0i}\psi_0(p+1) + \beta^R_{1i}\psi_1(p+1)+\beta^R_{2i}\psi(p+1). \tag{4.5.2}$$

The sums of squared differences between observed and predicted values for the 7-th measurement over the 13 mice for different methods using a second degree polynomial were as follows:

Method:	BLUP	JSRP	SRP	EBP	RRP
SSD:	.1042	.1044	.0972	.0951	.1047

In the problem considered, ridging (RRP) and James-Stein procedure (JSRP) do not seem to improve the least squares estimators (BLUP) for prediction purposes. As established theoretically in section 3, the EBP showed the best performance, while SRP is a close competitor.

It is proposed to study further methods of estimation of regression parameters and also transformations of the time axis to improve predictive efficiency.

BIBLIOGRAPHY

[1] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In 2nd International Symposium on Information Theory, Eds. B.N. Petrov and F. Csaki, pp. 267-81, Budapest, Akademia Kiado.

[2] Efron, B. and Morris, C. (1972). Empirical Bayes on vector observations. Biometrika 55, 335-47.

[3] Efron, B. and Morris, C. (1975). Data analysis using Stein's estimator and its generalizations. J. Am. Statist. Assoc. 70, 311-319.

[4] Hoel, A.E. And Kennard, R.W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12, 55-68.

[5] James, W. and Stein, C. (1961). Estimation with quadratic loss. Proc. 4th Berkeley Symp. 1, 362-379.

[6] Lindley, D.V. and Smith, A.F.M. (1972). Bayes estimates for the linear model (with discussion). J.R. Statist. Soc. B 34, 1-41.

[7] Rao, C. Radhakrishna (1953). Discriminant function for genetic differentiation and selection. Sankhyā 12, 229-246.

[8] Rao, C. Radhakrishna (1973). Linear Statistical Inference and its Applications, Wiley, New York.

[9] Rao, C. Radhakrishna (1975). Simultaneous estimation of parameters in different linear models and applications to biometric problems. Biometrics, 31, 545-554.

[10] Rao, C. Radhakrishna and Shinozaki, N. (1978). Precision of individual estimators in simultaneous estimation of parameters. Biometrika 65, 23-30.

[11] Rao, C. Radhakrishna (1980). Some comments on the minimum mean square error as a criterion of estimation. In Statistics and Related Topics, pp. 123-143 (Eds. Csörgö, Dawson, Rao and Saleh).

[12] Shibata, R. (1981). An optimal selection of regression variables. Biometrika 68, 45-54.

[13] Stein, C. (1955). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. Proc. 3rd Berkeley Symp. 1, 197-206.

[14] Toutenburgh, H. (1970). Vorhersage im allegemeinen linearen regression modell mit stochastishen regressoren. <u>Op. Forshung und Math. Stat</u>. 2, 105-116.

[15] Williams, J.S. and Izenman, A.J. (1981). A class of linear spectral models and analysis for the study of longitudinal data. Tech. Rept. Dept. of Stat., Colorado State University.

ATE

LMED

-8